

Subset Typicality Lemmas and Improved Achievable Regions in Multiterminal Source Coding

Kumar Viswanatha, Emrah Akyol and Kenneth Rose
ECE Department, University of California - Santa Barbara
{kumar,eakyol,rose}@ece.ucsb.edu

Abstract—Consider the following information theoretic setup wherein independent codebooks of N correlated random variables are generated according to their respective marginals. The problem of determining the conditions on the rates of codebooks to ensure the existence of at least one codeword tuple which is jointly typical with respect to a given joint density (called the multivariate covering lemma) has been studied fairly well and the associated rate regions have found applications in several source coding scenarios. However, several multiterminal source coding applications, such as the general multi-user Gray-Wyner network, require joint typicality *only* within subsets of codewords transmitted. Motivated by such applications, we ask ourselves the conditions on the rates to ensure the existence of at least one codeword tuple which is jointly typical within subsets according to given per subset joint densities. This report focuses primarily on deriving a new achievable rate region for this problem which strictly improves upon the direct extension of the multivariate covering lemma, which has quite popularly been used in several earlier work. Towards proving this result, we derive two important results called ‘subset typicality lemmas’ which can potentially have broader applicability in more general scenarios beyond what is considered in this report. We finally apply the results therein to derive a new achievable region for the general multi-user Gray-Wyner network.

Index Terms—Typicality within subsets, Multivariate covering lemma, Multi-user Gray-Wyner network

I. INTRODUCTION

Consider a scenario where independent codebooks of N random variables (X_1, X_2, \dots, X_N) are generated according to some given marginal distributions at rates (R_1, R_2, \dots, R_N) respectively. Let $S_1, S_2 \dots S_M$ be M subsets of $\{1, 2, \dots, N\}$ and let the joint distributions of (X_1, X_2, \dots, X_N) within each subset, consistent with each other and with the marginal distributions, be given. We ask ourselves the conditions on the rates (R_1, R_2, \dots, R_N) (achievable region) so that the probability of finding one codeword from each codebook, such that the codewords are all jointly typical within

subsets $S_1, S_2 \dots S_M$ according to the given per subset joint distributions, approaches 1. We denote the given probability distribution over subset S_i by $P(\{X\}_{S_i})$. The conditions on the rates when $S_i = \{1, \dots, N\}$, i.e, when the joint distribution over all the random variables is given, can be derived using standard typicality arguments and is quite popularly called as the multivariate covering lemma [1], [2]¹. It says that for any joint density over (X_1, X_2, \dots, X_N) , if the codebooks are generated according to the respective marginals, the probability of not finding a jointly typical codeword tuple approach 0 if $\forall \mathcal{J} \subseteq \{1, 2, \dots, N\}$:

$$\sum_{i \in \mathcal{J}} R_i \geq \sum_{i \in \mathcal{J}} H(X_i) - H(P(\{X\}_{\mathcal{J}})) \quad (1)$$

where $\{X\}_{\mathcal{J}}$ denotes the set $X_i : i \in \mathcal{J}$ and $H(P)$ denotes the entropy of any distribution P .

A fairly direct extension of the multivariate covering lemma, to the more general scenario of arbitrary subsets $S_1, S_2 \dots S_M$, which has been quite popularly used in several information theoretic scenarios, such as [2], [4], [5], [6], [7], can be described as follows. Fix any joint density $\tilde{P}(X_1, X_2, \dots, X_N)$ such that:

$$\tilde{P}(\{X\}_{S_j}) = P_{S_j}(\{X\}_{S_j}) \quad \forall j \quad (2)$$

i.e, it satisfies the given joint distributions within subsets $S_j \forall j$. Then the set of all rate tuples satisfying the following conditions are achievable, $\forall \mathcal{J} \subseteq \{1, 2, \dots, N\}$:

$$\sum_{i \in \mathcal{J}} R_i \geq \sum_{i \in \mathcal{J}} H(X_i) - H(\tilde{P}(\{X\}_{\mathcal{J}})) \quad (3)$$

The convex closure of all achievable rate tuples, over all such joint densities \tilde{P} satisfying the given per subset densities is an achievable region for the problem. We denote this region by \mathcal{R}_a . Our primary objective in

¹We note that the underlying principles and proofs of multivariate covering lemma appeared much earlier in the literature, for example [3]. However the nomenclature and the general applicability of the underlying ideas have been elucidated quite clearly in [1]

this report is to show that the rate region in (3) with the individual functionals set to their respective maxima subject only to their specific exact constraints is, in fact, achievable. Specifically we show that, each of the terms $H(\tilde{P}(\{X\}_{\mathcal{J}}))$ can be replaced with the corresponding maximum entropy functionals $H^*(\tilde{P}(\{X\}_{\mathcal{J}}))$ subject to only the constraints pertinent to subsets of $\{X\}_{\mathcal{J}}$. This allows us to achieve simultaneous optimum of all the functionals leading to a strictly larger achievable region than \mathcal{R}_a . Towards proving this result, we establish two important lemmas, namely ‘subset typicality lemmas’, which may prove to have much wider applicability in general scenarios beyond the scope of this report.

Scenarios depicted in the above example, where typicality within subsets of codewords is sufficient for decoding, arise quite frequently in several multiterminal source coding setups. One of the most typical examples is the multi-user generalization of the Gray-Wyner network [8] discussed in section III where the encoder observes K random variables and there are K sinks, each decoding one of the random variables upto a prescribed distortion constraint². The most general setting involves $2^K - 1$ branches (encoding rates), each being sent to a unique subset of the decoders. Observe that it is sufficient if all the codewords being sent to sink i are jointly typical with the i ’th source sequence and enforcing joint typicality of all the codewords in an unnecessary restriction. Similar settings arise in the context of dispersive information routing of correlated sources [7], fusion coding and selective retrieval in a database [6] and in several other scenarios which can be considered as particular cross-sections of the general L -channel ‘multiple descriptions’ (MD) problem [2], [4]. We note that, in this report, we demonstrate the workings of the underlying principle in the context of the example we described above. However it is important to note that the results we derive have implications in a wide variety of problems involving optimization of multiple functionals, each depending on a subset of the random variables, subject to constraints on their joint distributions.

II. MAIN RESULTS

In this section, we first establish the subset typicality lemmas which will finally lead to Theorems 1 and 2

²We note that [9] considers a particular generalization of the Gray-Wyner network to multiple users with applications in information theoretic security where a unique common branch is sent to all the decoders along with their respective individual rates. However we assert that the most general extension of the 2 user Gray-Wyner network will involve a combinatorial number of branches, each being sent to a unique subset of the decoders.

showing strictly larger achievable rates compared to \mathcal{R}_a . Throughout the report, we use the following notation. n independent and identically distributed (iid) copies of a random variable and its realizations are denoted by X_0^n and x_0^n respectively. Length n , ϵ -typical set of any random variable X , with distribution $P(X)$ is denoted³ by $\mathcal{T}_\epsilon^n(P(X))$. Throughout the report, for any set \mathcal{S} , we use the shorthand $\{U\}_{\mathcal{S}}$ to denote the set $\{U_i : i \in \mathcal{S}\}$. Note the difference between U_{123} , which is a single random variable and $\{U\}_{123}$, which is the set of random variables $\{U_1, U_2, U_3\}$. In the following Lemmas, we use the notation $P(A) \doteq 2^{-nR}$ to denote $2^{-n(R+\delta(\epsilon))} \leq P(A) \leq 2^{-n(R-\delta(\epsilon))}$ for some $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. To avoid resolvable but unnecessary complications, we further assume that there exists at least one joint distribution consistent with the prescribed per subset distributions for $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$.

A. Subset Typicality Lemmas

Lemma 1. Subset Typicality Lemma : Let (X_1, X_2, \dots, X_N) be N random variables taking values on arbitrary finite alphabets $(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N)$ respectively. Let their marginal distributions be $P_1(X_1), P_2(X_2), \dots, P_N(X_N)$ respectively. Let $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$ be M subsets of $\{1, 2, \dots, N\}$ and for all $j \in \{1, 2, \dots, M\}$, let $P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})$ be any given joint distribution for $\{X\}_{\mathcal{S}_j}$ consistent with each other and with the given marginal distributions. Generate sequences $x_1^n, x_2^n, \dots, x_N^n$, each independent of the other, where x_i^n is drawn iid according to the marginal distribution $P_i(X_i)$, i.e., $x_i^n \sim \prod_{l=1}^n P_i(x_{il})$. Then,

$$P\left(\{x\}_{\mathcal{S}_j}^n \in \mathcal{T}_\epsilon^n(P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j}))\right), \forall j \in \{1 \dots M\} \doteq 2^{-n(\sum_{i=1}^N H(X_i) - H(P^*))} \quad (4)$$

where P^* is a distribution over $(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N)$ which satisfies:

$$P^* = \arg \max_{\tilde{P}} H(\tilde{P}) \quad (5)$$

subject to $\tilde{P}(\{X\}_{\mathcal{S}_j}) = P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j}) \forall j \in \{1 \dots M\}$.

This Lemma essentially says that the total number of sequence tuples $(x_1^n, x_2^n, \dots, x_N^n)$ generated according to their respective marginals which are jointly ϵ -typical according to $P_{\mathcal{S}_i}(\{X\}_{\mathcal{S}_i}) \forall i$ within subsets $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$, is approximately $2^{nH(P^*)}$ where P^* is the **maximum entropy** distribution subject to the constraint that the joint density within subset \mathcal{S}_i is $P_{\mathcal{S}_i}(\{X\}_{\mathcal{S}_i}) \forall i$.

³The parenthesis is dropped whenever it is obvious

Proof: To prove this Lemma, we resort to Sanov's theorem ([10] Theorem 11.4.1) from the theory of large deviations. Sanov's theorem states that for any distribution $Q(X)$ and for any subset of probability distributions $\mathcal{E} \subseteq \mathcal{P}$, where \mathcal{P} denotes the universe of the PMFs over the alphabets of X :

$$Q^n(\mathcal{E}) \doteq 2^{-nD(P^*||Q)} \quad (6)$$

for sufficiently large n , where P^* is the distribution closest in relative entropy to Q in \mathcal{E} and $Q^n(\mathcal{E})$ denotes the probability that an iid sequence generated according to $Q(X)$ is ϵ -typical with respect to some distribution in \mathcal{E} . We set $Q(\cdot) = \prod_{i=1}^N P_i(X_i)$ and \mathcal{E} as the set of all distributions over $(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N)$ satisfying the given constraints. Then it follows from Sanov's theorem that the probability of $(x_1^n \dots x_N^n)$ being ϵ -typical according to some distribution satisfying the given constraints is approximately $2^{-nD(P^*||\prod_{i=1}^N P_i(X_i))}$, where P^* is the distribution having minimum relative entropy to $\prod_{i=1}^N P_i(X_i)$ and satisfying the given constraints. However, all such distributions have the same marginal distributions $P_i(X_1), P_i(X_2), \dots, P_i(X_N)$. Hence minimizing relative entropy is equivalent to maximizing the joint entropy leading to P^* as defined in (5). Therefore we have:

$$\begin{aligned} P(\{x\}_{\mathcal{S}_i}^n \in \mathcal{T}_\epsilon^n(\{X\}_{\mathcal{S}_i}), \forall i) &= Q^n(\mathcal{E}) \doteq 2^{-nD(P^*||Q)} \\ &\doteq 2^{-n(\sum_{i=1}^N H(X_i) - H(P^*))} \end{aligned} \quad (7)$$

where the last equality follows because P^* satisfies the given marginals. ■

We note that a particular instance of Lemma 1 was derived in [11]. However, as it turns out, for the setup they consider, this Lemma does not help in deriving an improved achievable region. In the following lemma, we establish the conditional version of Lemma 1. Note that Lemma 2 is not used in proving Theorems 1 or 2, but will play a crucial role in the application of these results to more general multi-terminal source coding scenarios (as we will see in section III).

Lemma 2. Conditional Subset Typicality Lemma : Let random variables (X_1, X_2, \dots, X_N) , sets $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$ and joint densities $P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})$ be defined as in Lemma 1. Let the sequences $(x_1^n \dots x_N^n)$ be generated such that each sequence is generated conditioned on a subset of already generated sequences $\{x\}_{\mathcal{A}_i}^n$ and independent of the rest, where $(i, \mathcal{A}_i) \in \mathcal{S}_j$ for some $j \in \{1, \dots, M\}$.

Then we have:

$$P(\{x\}_{\mathcal{S}_i}^n \in \mathcal{T}_\epsilon^n(\{X\}_{\mathcal{S}_i}) \forall i \in \{1 \dots M\}) \doteq 2^{-n(\sum_{i=1}^N H(X_i|\{X\}_{\mathcal{A}_i}) - H(P^*))} \quad (8)$$

where P^* satisfies (5).

Proof: The proof follows in very similar lines to that of Lemma 1 by setting $Q(\cdot) = \prod_{i=1}^N P(X_i|X_{\mathcal{A}_i})$, as conditioning on $x_{\mathcal{A}_i}^n$ only introduces further constraints, which are redundant, as $P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})$ are consistent with each other and $(i, \mathcal{A}_i) \in \mathcal{S}_j$ for some $j \in \{1, \dots, M\}$. ■

B. Simultaneous Optimality of Functionals

In this section we will show that simultaneous optimality of all function $H(\tilde{P}(\{X\}_{\mathcal{J}}))$ is in fact achievable leading to a new achievable rate region for the problem stated in the introduction.

Theorem 1. Let random variables (X_1, X_2, \dots, X_N) , sets $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$ and joint densities $P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})$ be defined as in Lemma 1. For each $i \in \{1, 2, \dots, M\}$, let $x_i^n(m_i)$ $m_i \in \{1, \dots, 2^{nR_i}\}$ be independent sequences drawn iid according to the respective marginals, i.e., $x_i^n(m_i) \sim \prod_{l=1}^n P_i(x_{il}(m_i)) \forall m_i \in \{1, \dots, 2^{nR_i}\}$. Then $\forall \epsilon > 0$, $\exists \delta(\epsilon)$ such that $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ and,

$$\begin{aligned} P\left(\{x\}_{\mathcal{S}_j}^n (\{m\}_{\mathcal{S}_j}) \in \mathcal{T}_\epsilon^n(P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})) \forall j \right. \\ \left. \text{for some } \{m_1, m_2, \dots, m_N\}\right) \geq 1 - \delta(\epsilon) \end{aligned} \quad (9)$$

if, (R_1, R_2, \dots, R_N) satisfy the following conditions $\forall \mathcal{J} \subseteq \{1, 2, \dots, N\}$:

$$\sum_{i \in \mathcal{J}} R_i \geq \sum_{i \in \mathcal{J}} H(X_i) - H^*(\{X\}_{\mathcal{J}}) + \epsilon \quad (10)$$

where,

$$H^*(\{X\}_{\mathcal{J}}) = \max_{\tilde{P}(\{X\}_{\mathcal{J}})} H(\tilde{P}(\{X\}_{\mathcal{J}})) \quad (11)$$

where $\tilde{P}(\{X\}_{\mathcal{J}})$ satisfies:

$$\tilde{P}(\{X\}_{\mathcal{J} \cap \mathcal{S}_j}) = P(\{X\}_{\mathcal{J} \cap \mathcal{S}_j}) \forall j \in \{1 \dots M\} \quad (12)$$

We denote the rate region in (10) by \mathcal{R}_a^* .

Remark 1. Note that $H^*(\{X\}_{\mathcal{J}}) = H(P(\{X\}_{\mathcal{J}}))$ if $\mathcal{J} \subseteq \mathcal{S}_j$ for some \mathcal{S}_j . Hence for all \mathcal{J} such that $\mathcal{J} \subseteq \mathcal{S}_j$ for some j , the corresponding inequalities in Theorem 1 and equations (2) are the same. However this theorem asserts that for every other \mathcal{J} , the functionals in (2) can be replaced with the 'maximum joint entropy' subject

to the given subset distributions which involve only the random variables $\{X\}_{\mathcal{J}}$. It is very important to note that the maximum entropy distributions for two different subsets $X_{\mathcal{J}_1}$ and $X_{\mathcal{J}_2}$, $\mathcal{J}_1, \mathcal{J}_2 \subseteq \{1, 2, \dots, N\}$, may not even correspond to any valid joint distribution over (X_1, X_2, \dots, X_N) . This is precisely what provides the additional leeway in achieving points which are strictly outside (2) as illustrated in Theorem 2. A pictorial representation of the above theorem is shown in Fig. 1.

Proof: We are interested in finding conditions on rates so that the probability in (9) approaches 1. Denote the event $\mathcal{E} = \{m_1, m_2, \dots, m_N\} : \{x\}_{\mathcal{S}_j}^n(\{m\}_{\mathcal{S}_j}) \in \mathcal{T}_\epsilon^n(P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})) \forall j$. We want to make $P(\mathcal{E}) \rightarrow 0$. Let \mathcal{N} denote the set $\{1, 2, \dots, N\}$ and let $(m_1, m_2, \dots, m_N) = \{m\}_{\mathcal{N}}$ be an index tuple, one from each codebook, such that $m_i \in \{1, \dots, 2^{nR_i}\}$. Let $\mathcal{E}(\{m\}_{\mathcal{N}})$ denote the event that $\{x\}_{\mathcal{S}_j}^n(\{m\}_{\mathcal{S}_j}) \in \mathcal{T}_\epsilon^n(P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})) \forall j$. Define random variables $\chi(\{m\}_{\mathcal{N}})$ such that:

$$\chi(\{m\}_{\mathcal{N}}) = \begin{cases} 1 & \text{if } \mathcal{E}(\{m\}_{\mathcal{N}}) \text{ occurs} \\ 0 & \text{else} \end{cases} \quad (13)$$

and random variable $\chi = \sum_{\{m\}_{\mathcal{N}}} \chi(\{m\}_{\mathcal{N}})$. Then we have $P(\mathcal{E}) = P(\chi = 0)$. From Chebyshev's inequality, it follows that:

$$\begin{aligned} P(\mathcal{E}) &= P(\chi = 0) \leq P[|\chi - E(\chi)| \geq E(\chi)/2] \quad (14) \\ &\leq \frac{4\text{Var}(\chi)}{(E(\chi))^2} = \frac{4(E(\chi^2) - (E(\chi))^2)}{(E(\chi))^2} \end{aligned}$$

We next bound $E(\chi)$ and $E(\chi^2)$ using Lemma 1. First we write $E(\chi)$ as:

$$E(\chi) = 2^{n \sum_{i=1}^N R_i} P(\mathcal{E}(\{m\}_{\mathcal{N}})) \quad (15)$$

for any $\{m\}_{\mathcal{N}}$ because all the sequences are drawn independent of each other. Next towards bounding $E(\chi^2)$, note that:

$$E(\chi^2) = \sum_{\{m\}_{\mathcal{N}}} \sum_{\{l\}_{\mathcal{N}}} P(\mathcal{E}(\{m\}_{\mathcal{N}}), \mathcal{E}(\{l\}_{\mathcal{N}})) \quad (16)$$

Let $\{m\}_{\mathcal{Q}} = \{l\}_{\mathcal{Q}}$ and $\{m\}_{\mathcal{N}-\mathcal{Q}} \neq \{l\}_{\mathcal{N}-\mathcal{Q}}$ for some $\mathcal{Q} \subseteq \mathcal{N}$, $\mathcal{Q} \neq \phi$ where ϕ denotes a null-set. Then,

$$\begin{aligned} P(\mathcal{E}(\{m\}_{\mathcal{N}}), \mathcal{E}(\{l\}_{\mathcal{N}})) &= \left\{ P(\mathcal{E}(\{m\}_{\mathcal{Q}})) \right. \\ &\quad \left. P(\mathcal{E}(\{m\}_{\mathcal{N}}) | \mathcal{E}(\{m\}_{\mathcal{Q}}))^2 \right\} \quad (17) \end{aligned}$$

where $\mathcal{E}(\{m\}_{\mathcal{Q}})$ denotes the event that $\{x\}_{\mathcal{S}_j \cap \mathcal{Q}}^n(\{m\}_{\mathcal{S}_j \cap \mathcal{Q}}) \in \mathcal{T}_\epsilon^n(P_{\mathcal{S}_j \cap \mathcal{Q}}(\{X\}_{\mathcal{S}_j \cap \mathcal{Q}})) \forall j$, as conditional on $\{x\}_{\mathcal{Q}}^n(\{m\}_{\mathcal{Q}})$, sequences $\{x\}_{\mathcal{N}-\mathcal{Q}}^n(\{m\}_{\mathcal{N}-\mathcal{Q}})$ and $\{x\}_{\mathcal{N}-\mathcal{Q}}^n(\{l\}_{\mathcal{N}-\mathcal{Q}})$ are drawn independently from the same distribution. The above expression can be rewritten as:

$$\begin{aligned} P(\mathcal{E}(\{m\}_{\mathcal{N}}), \mathcal{E}(\{l\}_{\mathcal{N}})) &= \left\{ P(\mathcal{E}(\{m\}_{\mathcal{Q}})) \right. \\ &\quad \left. \times \left(\frac{P(\mathcal{E}(\{m\}_{\mathcal{N}}))}{P(\mathcal{E}(\{m\}_{\mathcal{Q}}))} \right)^2 \right\} \quad (18) \end{aligned}$$

If $\mathcal{Q} = \phi$, we have $P(\mathcal{E}(\{m\}_{\mathcal{N}}), \mathcal{E}(\{l\}_{\mathcal{N}})) = (P(\mathcal{E}(\{m\}_{\mathcal{N}})))^2$. Hence, we can write $\text{Var}(\chi)$ as:

$$\begin{aligned} \text{Var}(\chi) &= \sum_{\mathcal{Q} \subseteq \mathcal{N}, \mathcal{Q} \neq \phi} \left\{ 2^{n \sum_{i \in \mathcal{Q}} R_i + 2n \sum_{i \in \mathcal{N}-\mathcal{Q}} R_i} \right. \\ &\quad \left. \times P(\mathcal{E}(\{m\}_{\mathcal{Q}})) \left(\frac{P(\mathcal{E}(\{m\}_{\mathcal{N}}))}{P(\mathcal{E}(\{m\}_{\mathcal{Q}}))} \right)^2 \right\} \quad (19) \end{aligned}$$

Note that the $\mathcal{Q} = \phi$ term gets cancelled with the $(E(\chi))^2$ terms in $\text{Var}(\chi)$ (see [2] for a similar argument).

On substituting (15) and (19) in (14), and noting that for any $\mathcal{Q} \subseteq \mathcal{N}$, $\mathcal{Q} \neq \phi$, we can write $P(\mathcal{E}(\{m\}_{\mathcal{N}})) = P(\mathcal{E}(\{m\}_{\mathcal{Q}})) \frac{P(\mathcal{E}(\{m\}_{\mathcal{N}}))}{P(\mathcal{E}(\{m\}_{\mathcal{Q}}))}$, we have:

$$P(\mathcal{E}) \leq 4 \sum_{\mathcal{Q} \subseteq \mathcal{N}, \mathcal{Q} \neq \phi} 2^{-n \sum_{i \in \mathcal{Q}} R_i} (P(\mathcal{E}(\{m\}_{\mathcal{Q}})))^{-1} \quad (20)$$

Next, invoking Lemma 1, we bound $P(\mathcal{E}(\{m\}_{\mathcal{Q}}))$ as:

$$P(\mathcal{E}(\{m\}_{\mathcal{Q}})) \geq 2^{-n(\sum_{i \in \mathcal{Q}} H(X_i) - H^*(\{X\}_{\mathcal{Q}})) - n\delta(\epsilon)} \quad (21)$$

On substituting (21) in (20), it follows that $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$ if R_i satisfy (10). ■

C. Strict Improvement

Theorem 2. (i) The region in Theorem 1 subsumes the region in (3). i.e.,

$$\mathcal{R}_a \subseteq \mathcal{R}_a^* \quad (22)$$

(ii) There exist scenarios for which the region in Theorem 1 can be strictly larger than the region in (3). i.e.,

$$\mathcal{R}_a^* \supset \mathcal{R}_a \quad (23)$$

Proof: The first half of the Theorem follows directly because $H^*(\{X\}_{\mathcal{J}}) \geq H(\{X\}_{\mathcal{J}}) \forall \mathcal{J}$ for any joint distribution satisfying the given distributions within subsets. To prove (ii) we provide an example for which \mathcal{R}_a^* has points which are not part of \mathcal{R}_a . Consider the following

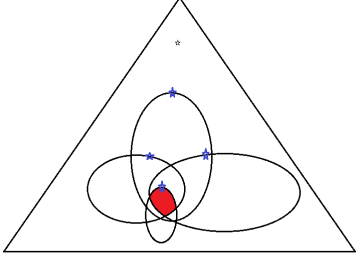


Figure 1. Pictorial representation of Theorem 1: The triangle denotes the simplex of all joint distributions over (X_1, X_2, \dots, X_N) . The black star denotes the joint distribution representing the product of marginals (codebook generation). Each loop represent the set of all joint distributions satisfying the conditions imposed on $\{X\}_{\mathcal{J}}$ for some \mathcal{J} . The intersection of all the loops (red region) represents the set of joint distributions satisfying all the conditions. The blue stars represent the joint distributions which maximize functionals $H(P(\{X\}_{\mathcal{J}}))$ (equivalently, minimize the relative entropy with the product of marginals as seen from Sanov's theorem) subject to the conditions on $\{X\}_{\mathcal{J}}$. Theorem 1 asserts that a separate joint distribution for each \mathcal{J} can be chosen from the corresponding loop (blue stars) and hence all the functionals $H(P(\{X\}_{\mathcal{J}}))$ can be set to their respective maxima simultaneously.

(x_i, x_4)	0, 0	0, 1	1, 0	1, 1
$P(x_i, x_4)$	$1/2$	0	$1/4$	$1/4$

Table I
PAIRWISE PMF OF $(X_i, X_4) \forall i \in \{1, 2, 3\}$

example of 4 binary random variables (X_1, X_2, X_3, X_4) . X_1, X_2 and X_3 are distributed $\text{bern}(\frac{1}{2})$ and X_4 is distributed $\text{bern}(\frac{3}{4})$, where $\text{bern}(p)$ denotes a Bernoulli random variable with $P(0) = p$ and $P(1) = 1 - p$. Let $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_6$ be all possible subsets of $\{1, 2, 3, 4\}$ of cardinality 2. Let $P_{\mathcal{S}_j}(\{X\}_{\mathcal{S}_j})$ be such that (X_1, X_2, X_3) are pairwise independent and the pairwise PMF of $(X_i, X_4) \forall i \in \{1, 2, 3\}$ is given in Table I. Note that these pairwise densities are satisfied by at least one joint density obtained by the following operations : $X_3 = X_1 \oplus X_2$ and $X_4 = X_1 \bullet X_2$, where X_1 and X_2 are independent $\text{bern}(\frac{1}{2})$ random variables and ' \oplus ' and ' \bullet ' denote 'bit-exor' and 'bit-and' operations respectively.

Observe that maximizing the entropy over (X_1, X_2, X_3) subject to their respective pairwise densities makes them mutually independent. However, there exists *no* joint distribution over (X_1, X_2, X_3, X_4) satisfying all the pairwise conditions which makes (X_1, X_2, X_3) mutually independent. This intuition is in fact sufficient to see that $\mathcal{R}_a^* \supset \mathcal{R}_a$. However to be more rigorous, we first rewrite the achievable region

\mathcal{R}_a^* for this example as:

$$\begin{aligned} R_i + R_4 &\geq H_b(\frac{1}{4}) - \frac{1}{2}H_b(\frac{1}{2}) \\ R_i + R_j + R_4 &\geq 2 + H_b(\frac{1}{4}) - H^*(X_i, X_j, X_4) \\ \sum_{i=1}^4 R_i &\geq 3 + H_b(\frac{1}{4}) - H^*(\{X\}_{1,2,3,4}) \end{aligned} \quad (24)$$

$\forall i, j \in \{1, 2, 3\}$ where $H_b(\cdot)$ denotes the binary entropy function and $\{X\}_{1,2,3,4} = \{X_1, X_2, X_3, X_4\}$.

We consider the following corner point of (24), $A = (0, 0, 0, 3 + H_b(\frac{1}{4}) - H^*(\{X\}_{1,2,3,4}))$. It is sufficient for us to prove that $A \notin \mathcal{R}_a$. Note that, if $R_1 = R_2 = R_3 = 0$, (X_1, X_2, X_3) must be mutually independent (which in turn satisfies the pairwise independence conditions). To prove that $A \notin \mathcal{R}_a$, we will show that there cannot exist *any* joint PMF over (X_1, X_2, X_3, X_4) satisfying all pairwise distributions and for which (X_1, X_2, X_3) are mutually independent. Let us suppose that such a joint PMF exists. Denote the conditional PMF $P(X_4 = 0 | x_1, x_2, x_3) = \alpha_{x_1 x_2 x_3}$, $x_1, x_2, x_3 \in \{0, 1\}$. As (X_1, X_2, X_3) are assumed to be mutually independent, the joint distribution $P_{X_1, X_2, X_3, X_4}(x_1, x_2, x_3, 1) = \frac{1 - \alpha_{x_1 x_2 x_3}}{8}$. The pairwise distribution of (X_1, X_4) (from Table I) is such that $P_{X_i, X_4}(0, 1) = 0 \forall i \in \{1, 2, 3\}$. This leads to the conclusion that $\alpha_{x_1 x_2 x_3} = 1$ if any one of x_1, x_2, x_3 is 0. We are only left with finding α_{111} . Further, we want $P_{X_1, X_4}(1, 1) = \frac{1}{4}$, i.e. $\sum_{x_2, x_3} P_{X_1, X_2, X_3, X_4}(1, x_2, x_3, 1) = \sum_{x_2, x_3} \frac{1 - \alpha_{1 x_2 x_3}}{8} = \frac{1}{4}$. One substituting, we have $\alpha_{111} = 2$. As α_{x_1, x_2, x_3} s are conditional probabilities, this leads to a contradiction and proves that there cannot exist a joint distribution with (X_1, X_2, X_3) being mutually independent. Therefore $\mathcal{R}_a^* \supset \mathcal{R}_a$, proving the second half of the Theorem. ■

III. APPLICATION TO MULTI-USER GRAY-WYNER NETWORK

We finally apply the results in Theorem 1 to obtain a new achievable region for the multi-user Gray-Wyner network. To illustrate the applicability and to maintain simplicity in notation, we only consider the 3-user lossless Gray-Wyner network here. However the approach can be extended directly to the general L -user setting and to incorporate distortions. Note that the formal definition of an achievable rate region closely resembles that in [8], with obvious generalization to the 3 user setting as shown in Fig. 2. We omit the details here due to space constraints. We further note that the rate region

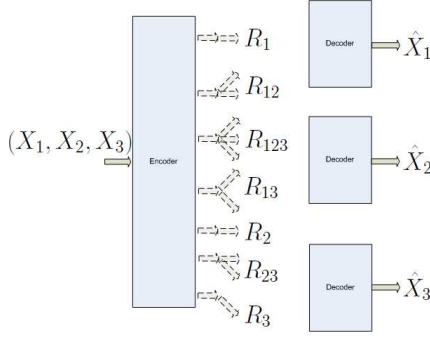


Figure 2. 3-user Gray-Wyner network: There is a unique branch from the encoder to every subset of the decoders

is in general 7 dimensional, with the following rates: $(R_1, R_2, R_3, R_{12}, R_{13}, R_{23}, R_{123})$.

Corollary 1. *Let (X_1, X_2, X_3) be the random variables with joint distribution $P(X_1, X_2, X_3)$ observed by the encoder. Let $(U_{123}, U_{12}, U_{13}, U_{23})$ be random variables jointly distributed with (X_1, X_2, X_3) with conditional distribution $P(U_{123}, U_{12}, U_{13}, U_{23}|X_1, X_2, X_3)$ and taking values over arbitrary finite alphabets. Define subsets $\mathcal{S}_1 = \{U_{123}, U_{12}, U_{13}\}$, $\mathcal{S}_2 = \{U_{123}, U_{12}, U_{23}\}$, $\mathcal{S}_3 = \{U_{123}, U_{13}, U_{23}\}$. The rate region for the 3-user lossless Gray-Wyner network contains all the rates such that $\forall (i, j, k) \in \{1, 2, 3\}$ and $i < j, i < k$,*

$$\begin{aligned}
R_{123} &\geq H(U_{123}) - H^*(U_{123}|\mathbf{X}) \\
R_{123} + R_{ij} &\geq H(U_{123}, U_{ij}) \\
&\quad - H^*(U_{123}, U_{ij}|\mathbf{X}) \\
R_{123} + R_{ij} + R_{ik} &\geq H(U_{123}) - H^*(\{U\}_{123,ij,ik}|\mathbf{X}) \\
&\quad + H(U_{ij}|U_{123}) + H(U_{ik}|U_{123}) \\
R_{123} + \sum_{i < j} R_{ij} &\geq H(U_{123}) + \sum_{i < j} H(U_{ij}|U_{123}) \\
&\quad - H^*(U_{123}, U_{12}, U_{23}, U_{13}|\mathbf{X}) \\
R_i &\geq H(X_i|\{U\}_{\mathcal{J}: i \in \mathcal{J}}) \quad (25)
\end{aligned}$$

where $\mathbf{X} = \{X_1, X_2, X_3\}$ and $H^*(\{U\}_{\mathcal{J}}|\mathbf{X})$ is given by:

$$\max_{\tilde{P}(\{U\}_{\mathcal{J}}, \{X\}_{1,2,3})} H(\tilde{P}(\{U\}_{\mathcal{J}}|\mathbf{X})) \quad (26)$$

where $\tilde{P}(\{U\}_{\mathcal{J}}|\mathbf{X})$ satisfies:

$$\tilde{P}(\{U\}_{\mathcal{J} \cap \mathcal{S}_j}, X_j) = P(\{U\}_{\mathcal{J} \cap \mathcal{S}_j}, X_j) \quad \forall j \quad (27)$$

The closure of the achievable rates over all conditional distributions $P(U_{123}, U_{12}, U_{13}, U_{23}|X_1, X_2, X_3)$ is an achievable region for the 3-user lossless Gray-Wyner network.

Proof: A codebook for U_{123} consisting of $2^{nR_{123}}$ codewords is generated according to the marginal $P(U_{123})$. Conditioned on each codeword of U_{123} , independent codebooks are generated for U_{12}, U_{13} and U_{23} at rates of R_{12}, R_{13} and R_{23} according to their respective conditional distributions $P(U_{12}|U_{123})$, $P(U_{13}|U_{123})$ and $P(U_{23}|U_{123})$. If the rates satisfy (25), then there always exists a codeword tuple, one from each codebook, denoted by $(u_{123}^n, u_{12}^n, u_{13}^n, u_{23}^n)$, such that the following subsets of sequences are jointly typical according to their respective subset joint densities: $(x_1^n, u_{123}^n, u_{12}^n, u_{13}^n)$, $(x_2^n, u_{123}^n, u_{12}^n, u_{23}^n)$ and $(x_3^n, u_{123}^n, u_{13}^n, u_{23}^n)$. The proof follows rather directly from Lemmas 1, 2 and Theorem 1 as U_{123} is part of $\mathcal{S}_1, \mathcal{S}_2$ and \mathcal{S}_3 . The last constraint in (25) denotes the minimum rate of the bin indices required to achieve lossless reconstruction at each sink given that all the codewords received at any sink are jointly typical. ■

IV. DISCUSSION

We note that the conditions in (25) ensure joint typicality of source sequence X_i^n only with the codewords which reach sink i . However an alternate achievable region (which is subsumed in the above region) can be derived using results of the general L -channel MD problem in [2] which extends the principles underlying (3) to the multiple descriptions framework. Due to the inherent structure of the MD problem, joint typicality of all the transmitted codewords is necessary. However imposing such a constraint limits the performance of systems that do not explicitly require such conditions. Note that, although we have not proved formally that the new region for the multi-user Gray-Wyner network is strictly larger than that derivable from the results in [2], Theorem 2 suggests that for general sources, there exist points which are strictly outside. It is important to note that implications of the results we derived may not always lead to a strictly larger achievable region. A classic example of this setting is the 2 user Gray-Wyner network [8] for which the complete rate-distortion region can be achieved even if joint typicality of all the codewords is imposed. This is because, in the 2-user scenario, there is no inherent conflict between maximum entropy distributions of different subsets of random variables. However, in the L -user setting (as seen in Theorem 2), such a conflict arises and maintaining joint typicality only within subsets plays a paramount role in deriving improved achievable regions.

REFERENCES

- [1] A. El-Gamal, Y.H. Kim, "Lecture notes on network information theory", 23-61 to 23-67, <http://arxiv.org/abs/1001.3404>, 2010.
- [2] R. Venkataramani, G. Kramer, V.K. Goyal, "Multiple description coding with many channels", IEEE Trans. on Information Theory, vol.49, no.9, pp. 2106- 2114, Sept 2003.
- [3] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," IEEE Trans. Inf. Theory, vol. IT-28, pp. 851–857, Nov. 1982.
- [4] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptions-part II: an achievable rate-distortion region", IEEE Trans. Information Theory, vol. 51, pp. 1377-1392, Apr. 2005.
- [5] K. Viswanatha, E. Akyol and K. Rose, "Combinatorial message sharing for a refined multiple descriptions achievable region", in Proc. IEEE Symp. Information Theory (ISIT), Aug. 2011.
- [6] J. Nayak, S. Ramaswamy, K. Rose, "Correlated source coding for fusion storage and selective retrieval", in Proc. IEEE Symp. Information Theory (ISIT), Sept. 2005.
- [7] K. Viswanatha, E. Akyol and K. Rose, "An achievable rate region for distributed source coding and dispersive information routing" in Proc. IEEE Symp. Information Theory (ISIT), Aug. 2011.
- [8] R. Gray and A. Wyner, "Source coding for a simple network", Bell systems technical report, Dec 1974.
- [9] R. Tandon, L. Sankar, and H. V. Poor, "Multi-user privacy: The Gray-Wyner system and generalized common information," in Proc. IEEE Symp. Information Theory (ISIT), Aug. 2011.
- [10] T. Cover and J. Thomas, "Elements of Information Theory", Wiley publications, Second edition, 2006.
- [11] E. Perron, S. Diggavi, E. Telatar, "On the role of encoder side-information in source coding for multiple decoders," In Proc. IEEE International Symposium on Information Theory (ISIT), vol., no., pp.331-335, 9-14 Jul 2006.